14A.2

THE USE OF A MODIFIED EBERT-MCBRIDE TECHNIQUE TO EVALUATE MESOSCALE MODEL QPF AS A FUNCTION OF CONVECTIVE SYSTEM MORPHOLOGY DURING IHOP 2002

Jeremy S. Grams*, William A. Gallus, Jr.^, Steven E. Koch[+], Linda S. Wharton[+], Andrew Loughe[+], and Elizabeth E. Ebert[#]

*NOAA/NWS Meteorological Development Laboratory, Silver Spring, Maryland
^Iowa State University, Ames, Iowa
[+]NOAA/OAR Forecast Systems Laboratory, Boulder, Colorado
[#]Bureau of Meteorology Research Center, Melbourne, Victoria, Australia

1. INTRODUCTION

Verification of quantitative precipitation forecasts (QPFs) made by fine-grid numerical models for mesoscale convective systems (MCSs) can be problematic. Traditional verification statistics severely penalize a precipitation system that may have been forecast with a small positional error or incorrect shape, with resultant low correlation coefficients, high root mean square errors (RMSEs), and poor values of categorical statistics (Ebert and McBride 2000; Baldwin and Wandishin 2002). Numerous approaches have been applied to deal with the deficiencies of traditional verification methods (e.g., Du et al. 2000; Zepeda-Arce et al. 2000; Bullock et al. 2004). One such approach is the Ebert-McBride technique (EMT), which employs the concept of matching individual forecast and observed areas (Ebert and McBride 2000, hereafter EM2000). The technique utilizes contiguous rain areas (CRAs), defined as the areas of contiguous observed and forecast rainfall enclosed within a specified isohyet. A displacement is performed by an objective pattern matching technique to optimally align the forecast with the observations. The EMT method was originally applied to synoptic-scale precipitation systems over Australia. The current study adapts the EMT to warm-season MCSs occurring over the central U.S.

The International H_2O Project (IHOP) that took place from 16 May to 26 June 2002 was designed to help improve the understanding and prediction of QPF. High-resolution model datasets produced for this project offered the opportunity to investigate precipitation forecast accuracy as a function of convective system morphology. Analysis from the EMT objective verification measures in concert with an observed morphological classification scheme revealed systematic errors for certain types of MCSs.

2. EMT BACKGROUND

The EMT was used to evaluate the performance of three 12-km models: the NCEP operational Eta model (Rogers et al. 1998) with Betts-Miller-Janjic (BMJ) convective parameterization (Janjic 1994), the Penn State/NCAR MM5 model version 3.5 (Grell et al. 1995), and the WRF model version 1.3 (Skamarock et al. 2001). Both the MM5 and WRF were run with no convective parameterization and initialized with a “Hot Start” procedure (McGinley and Smart 2001) developed for the NOAA Forecast Systems Laboratory, Local Analysis and Prediction System (Albers et al. 1996).

2.1 Overview of EMT

The EMT uses CRAs as a way to determine error statistics. These CRAs are made up of the union of observed and forecast rainfall areas which exceed a user-specified threshold amount. An optimal displacement vector is then determined by translating the forecast area over the observed area, typically by either maximizing correlation coefficient or by minimizing the total squared error. The forecast is permitted to shift within an expanded box enclosing the CRA (the maximum distance allowed between the forecast and observed areas beyond which it is assumed the two areas are unrelated). Several user-defined parameters can be adjusted to define the temporal and spatial scale of the CRA, the pattern matching process, and how verification statistics are calculated.

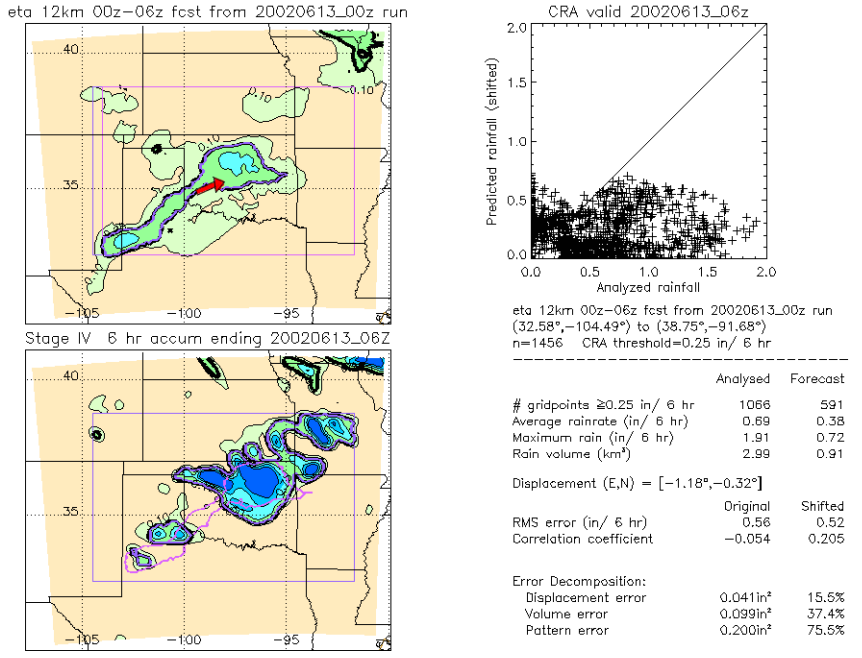
Figure 1 shows two examples of CRA output from 0000 UTC 13 June 2002 for the a) Eta and b) WRF 6-h forecast of precipitation, respectively, in the upper left of each plot; with the smoothed NCEP stage IV 6-h accumulated precipitation product in the bottom left panel. A displacement vector (in red) is determined by shifting the forecast entity to maximize the correlation coefficient between the forecast and observed entities. Various measures of error were determined before and after displacement (shown to the right in Fig. 1).

2.2 Improvements to the EMT

The EMT objectifies the intuitive process of pattern matching. It is therefore important to choose values of parameters that give the best agreement between the objective pattern matching and the investigator's visual interpretation. One of the advantages of this technique is that a variety of arbitrary parameters can be tuned based on the needs of the user. The following sections

* *Corresponding author address:* Jeremy S. Grams, NOAA/NWS Meteorological Development Laboratory, 1325 East-West Highway, Silver Spring, MD 20910; e-mail: jeremy.grams@noaa.gov

a)



b)

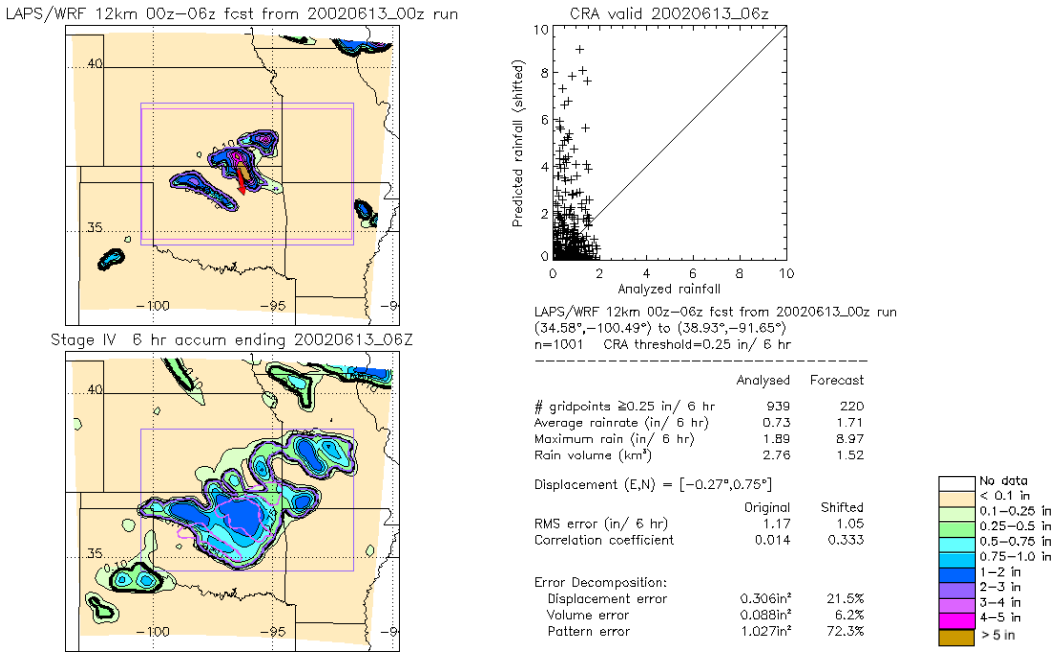


Figure 1. Example of CRA output from the Ebert-McBride technique for the 0000 UTC run of the 12-km a) Eta and b) WRF models on 13 June 2002. In both (a) and (b), the 6-h model forecast of rain above the 0.25 inch threshold is outlined in purple in the upper left. Displacement vectors (in red) show computed displacement of forecast rain area to the northeast for the Eta and to the south-southeast for the WRF. Stage IV 6-h observed rainfall accumulation above the 0.25 inch threshold is outlined in purple in the lower left of both (a) and (b), with the shifted forecast overlaid in magenta. Outer purple box shows the area over which CRA statistics (shown to the right) were calculated. The graph in the upper right of both (a) and (b) shows point-to-point verification of the shifted forecast rainfall versus observed rainfall. The tables in the center right of both (a) and (b) show various statistical measures used in the study. The legend in the lower right of (b) shows the thresholds for the 6-h rainfall accumulations.

describe the modifications made for the purpose of verifying central U.S. MCSs.

a. CRA rainfall threshold

For 24-h QPF verification, EM2000 used a CRA critical rainfall threshold of 5 mm (~0.20 inches) per day for the minimum accumulation required for a grid point to be considered part of a CRA. For our purposes, a critical threshold of 0.25 inches in 6 hours was found to work reasonably well at corresponding to small MCSs over the central U.S.

This threshold was the most important element for the inclusion or division of multiple objects in a CRA. Since pattern matching generally requires both a forecast and observed entity, CRAs are model dependent. This can make it difficult to individually compare statistical results from different models. Gallus (1999) found that the Eta model using the BMJ convective scheme often depicted relatively large areas of contiguous low-to-moderate rainfall because of the design of the scheme. These smoothed patterns do not resemble typical observed rainfall patterns during warm season convective episodes. An example of this can be found by comparing Figs. 1a and 1b. One can see the broad area of low-to-moderate rainfall forecast in the Eta versus the intense, but small, area of rainfall forecast in the WRF. An overly broad forecast rainfall area can be responsible for two or more distinct observed systems getting combined into one large CRA.

b. Critical mass threshold

The critical mass threshold defines a minimum volume of rainfall necessary for a system to be identified by the EMT. Since our study focused on the first 6 hours of a model forecast, we chose a critical mass threshold (~3 x 10^{11} kg) corresponding to a combined forecast and observed system producing a minimum of 0.25 inches of rain in 6 hours over a 40000 km^2 area.

In the complete absence of a forecast system, the threshold will allow the EMT to identify observed systems whose spatial scales and intensities match the minimum radar-based criteria for a MCS. Every observed system from the primary MCS morphological types (discussed in section 3) were matched to a corresponding forecast. Systems with very little or no forecast rainfall but enough observed rainfall to meet the CRA critical mass threshold were included in the statistical analysis and were classified based on radar morphology just like any other CRA. Thus, the full spectrum of model forecasts to observed events was represented.

c. Search radius

The search radius allows for initially separated forecast and observed entities to be matched provided they are located within the search radius limit. After matching, these two rain areas become a single contiguous area. We chose a search domain of 20 grid points (240 km) over which a forecast system could be shifted to match an observed one. This was roughly equal to the length scale used in defining the critical mass threshold.

d. Pattern matching criterion

Hoffman et al. (1995) found that minimization of RMSE and maximization of correlation coefficient were the best methods for determining the fit of a forecast pattern to an observed one over a rectangular domain. EM2000 found that minimizing the total squared error gave the best pattern matches for 24-h QPFs, although they noted that maximizing the correlation coefficient generally gave similar results. In our study, maximization of correlation coefficient and minimization of total squared error generally gave similar displacements for most CRAs, agreeing with EM2000's findings. However, our tests showed that maximization of correlation coefficient worked better overall near the edge of the IHOP domain. When using total squared error minimization, forecast rain areas would typically shift off the verification grid instead of matching up with nearby observed systems. This type of shifting would result in the lowest total squared error calculation, by eliminating half of the double penalty (rain in the wrong place, no rain in the right place).

Maximizing correlation coefficient resulted in more reasonable matches and fewer problems of systems being shifted off of the domain. This matching strategy allowed the forecast rainfall maxima to be shifted to closely align with observed maxima, since correlation coefficient maximization matches rainfall gradients. In addition, for most cases, the use of correlation coefficient maximization resulted in little relative increase in total squared error. However, use of total squared error minimization resulted in much lower correlation coefficients for smaller CRAs.

e. Error decomposition

Forecast errors in rain events can be expressed in terms of errors in displacement, intensity, and pattern or variability of the rainfall (EM2000). The switch to correlation coefficient maximization instead of total squared error minimization was found to occasionally result in incorrect residual calculations of displacement errors that permitted negative RMSEs to occur. Therefore, a new error decomposition method was developed using correlation coefficient and mean square error (MSE) terms based on Murphy (1995). In that paper, MSE was represented as:

$$MSE = (\bar{g} - \bar{y})^2 + (s_g - r_o s_y)^2 + (1 - r_o^2) s_y^2 \quad (1)$$

where s represents the standard deviation, and r_o is the original correlation coefficient between the forecast (represented by y) and observed (represented by g) rain fields before the forecast is shifted by the EMT. Rearranging the second and third terms gives:

$$MSE = (\bar{g} - \bar{y})^2 + (s_g - s_y)^2 + 2s_g s_y (1 - r_o) \quad (2)$$

The first term is the unconditional bias, or volume error (MSE_{volume}). The second term compares the sample standard deviations of the forecast and observations and is a type of pattern error ($MSE_{pattern}$). The third term contains additional pattern error and the displacement error. These can be separated by adding and subtracting *r* (optimal correlation) in the third term:

$$MSE = (\bar{g} - \bar{y})^2 + (s_g - s_y)^2 + 2s_g s_y (1 - r) + 2s_g s_y (r - r_o) \quad (3)$$

The third term in (3) represents the shape, or fine scale pattern error ($MSE_{pattern}$), since it includes the difference between a perfect correlation (r=1) and the optimal correlation for the forecast, *r*. The fourth term in (3) represents the contribution of displacement error ($MSE_{displacement}$), as it includes the difference in covariances before and after shifting the forecast. Combining both the second and third terms in (3), the error decomposition (shown in Fig. 1) can be summarized in eq. (4) as:

$$MSE_{total} = MSE_{volume} + MSE_{pattern} + MSE_{displacement} \quad (4)$$

The error decompositions based on total squared error minimization (EM2000) and correlation maximization (eq. 3) produced very similar results. CRA verification of several thousand 24-h QPFs over Australia with both approaches gave mean pattern errors that were virtually identical, and differences of only a few percent between methods for volume and displacement errors (EM2000).

f. Verification statistics computation

Usually categorical statistics are computed over entire model domains. For an object-oriented technique, there is uncertainty over which areas should be used for calculating various verification statistics. In this study, four verification categories were adjusted. Rain volume, maximum rainfall, average rain rate, and number of gridpoints exceeding the user-defined threshold were previously calculated over the union of the observed, original forecast, and shifted forecast regions (EM2000). To better describe the characteristics of each individual entity, these parameters were computed exclusively over the observed and original forecast portions of the CRA before any displacement occurs. Only gridpoints at or above the CRA rainfall threshold were included in the analysis area for each portion (the areas enclosed by the purple isohyet in Fig. 1).

g. Filtering

It is well-known that models cannot adequately predict the spatial structure of small scales due to interpolation from finite differencing schemes and parameterized horizontal diffusive processes. The minimum resolvable feature varies as a function of not only the grid-spacing of models, but also the numerics and physics in each type of model. Most mesoscale models will generally be able to resolve only rainfall features of wavelength at roughly five times the grid spacing. Harris et al. (2001) showed that the 3-km ARPS model could not resolve less than 5 delta waves. Baldwin and Wandishin (2002) also found 3-5 delta waves to be the smallest resolvable wavelength in the 22-km Eta with the KF parameterization and in 10- and 22-km versions of the WRF model. However in the 12-km Eta with the BMJ parameterization, features less than 200 km were not resolved well, which might argue for filtering of 17 delta waves. In the present study, we decided that the stage IV observations should be filtered so that the observed rain areas resembled what the majority of the 10-12 km grid spacing models run by FSL during IHOP were able to show (Koch et al. 2004). Thus, the stage IV data were remapped to each native model grid and filtered using a low-pass Lanczos filter (Duchon 1979) to remove wavelengths less than 6 delta (72 km). This procedure does not remove any mismatch in the variability of the model QPFs. As will be shown in section 5e, error measures reflect the Eta's low variability in QPFs compared to the MM5 and WRF.

3. CLASSIFICATION OF CONVECTIVE SYSTEMS

A detailed radar-based morphological analysis of observed systems was performed for all CRAs identified in the IHOP domain. This was done for the first 6 hours of each model run available during the IHOP period. The observed system highlighted in the stage IV 6-h accumulated precipitation product was cross-referenced with an observed system indicated in radar observations. The radar-based morphology used 2-km NEXRAD composite base reflectivity radar imagery with a temporal resolution of 30 minutes.

We defined a radar-based MCS as a convective system containing continuous or discontinuous convective echoes that propagated and/or organized in nearly the same manner as other convective echoes within the system. We required the minimum MCS criteria to have at least 30 dBZ of base radar reflectivity over at least a 10000 km^2 (i.e., 100 x 100 km) area and at least 40 dBZ in a 2500 km^2 (i.e., 50 x 50 km) area. Both dBZ conditions had to exhibit temporal continuity for at least 3 hours. Using the Z-R relationships of $Z = 200 * R^{1.6}$ for stratiform and $Z = 300 * R^{1.4}$ for convective rain, 30 dBZ corresponds to a rain rate of around 0.10 in/hr with 40 dBZ corresponding to a rain rate of nearly 0.50 in/hr. The following sub-sections describe the definitions used for the observed system classification.

3.1 *General classification*

The first series of classifications began with a distinction between linear and non-linear systems for those meeting the MCS criteria. Since not all observed systems identified by the EMT met our radar-based MCS criteria, separate categories had to be made for these "marginal" systems. The classification scheme included seven general types of systems:

a. Continuous Linear (CL)

A continuous major axis of at least 40 dBZ convective echoes, of at least 100-km length, which share a nearly common leading edge and move approximately in tandem. In addition, the major axis must be at least three times as long as the minor axis.

b. Continuous Linear Bowing (CLB)

In addition to the linear criteria above, a linear bowing system must contain a bulging, convex shape (angle greater than 30°) of continuous convective cells with a tight reflectivity gradient on the front edge of the convective region. This shape must be identified for at least 1.5 hours on radar.

c. Continuous Non-Linear (CNL)

If the criteria for a CL or CLB were not met, but the system contains a contiguous region of echoes that satisfy the minimum size criteria for an MCS, then the CNL classification is given.

d. Discontinuous Areal (DA)

If the above minimum size MCS requirements were not met in a continuous area but were met in an area of discrete convective elements, in which no element is separated by more than 200 km from another, then the DA classification is given.

e. Isolated Cells (IC)

If discrete cells were too small, isolated, or lack temporal continuity to meet the DA classification, but had at least 40 dBZ in a 400 km^2 area and at least 30 dBZ in a 1600 km^2 region, then the IC classification is given. It is well-understood that a 12-km model cannot fully resolve isolated cell events.

f. Orographically Fixed (OF)

If a system remained nearly stationary with respect to the western edges of the IHOP domain (the Rocky Mountains and Black Hills), then the system was classified as OF since the mesoscale processes influencing these mountain systems may differ from systems over the Plains.

g. False Alarm (FA)

If none of the above criteria are met, then the observed system in the CRA is classified as a FA.

3.2 *Additional Linear Classifications*

For every linear type system (either CL or CLB), additional sets of classifications were performed by using the taxonomy proposed by Parker and Johnson (2000) and Bluestein and Jain (1985). First, the arrangement of stratiform rainfall with respect to the convective region was classified according to definitions given by Parker and Johnson (2000). Second, a classification was made based on Bluestein and Jain's (1985) four definitions for squall line development.

a. Stratiform Classification

Parker and Johnson (2000) defined three areas where stratiform precipitation was present with respect to convective precipitation in an MCS. Their three categories: trailing (TS), leading (LS), and parallel (PS), were used in this study. Combinations of these types were noted, when both were seen for at least 1.5 hours.

b. Development Classification

Bluestein and Jain (1985) defined four types of development for a squall line MCS. Their four categories: broken areal (BA), broken line (BL), back building (BB), and embedded areal (EA) were used in this study.

In the 6-h period over which CRAs were defined from accumulated rainfall data, multiple radar-based systems might be observed within one larger CRA. When this was the case, the system with the greater temporal, spatial, and/or rain volume was used to define the morphology of the CRA. In other cases, when the morphology of a single system changed over time, the morphology that occurred over the majority of the 6-h period was used to classify the CRA. It is understood that defining a single convective morphology for multiple radar-based systems will increase the amount of statistical uncertainty. However, it is pertinent to include these CRAs in the statistical analysis, since a clearly dominating type occurred in the vast majority of these cases.

4. OBSERVED MCS TYPE DISTRIBUTIONS

A total of 190 CRAs were identified for the Eta, 164 for the MM5, and 163 for the WRF during the IHOP period. Of the CRAs identified, 7% of Eta systems and 2% of MM5 and WRF systems were classified as FA (little or no observed rain); 4-5% were classified as OF to the Rockies and Black Hills at the western edge of the IHOP domain. IC systems accounted for 12% of the CRAs in the Eta, 6% in the MM5 and 5% in the WRF. Other than to note the number of occurrences, we exclude IC, FA, and OF systems (22% in the Eta, 13% in the MM5, and 12% in the WRF) from further analysis in this study. The IC and FA systems were only identified because of forecasted rainfall; observations did not show enough rain volume to meet the CRA critical mass threshold. The focus of subsequent evaluation is on model performance as a function of the observed system morphology of the 148 remaining events for the Eta, 144 events for the MM5, and 143 events for the WRF.

Fifty-five (37%) observed cases were classified as linear in the Eta, 62 (43%) in the MM5, and 60 (42%) in the WRF. Ninety-three (63%) observed cases were classified as non-linear in the Eta, 82 (57%) in the MM5,

and 83 (58%) in the WRF. Figure 2 shows a histogram of general, squall, and development types for every identified CRA which met MCS criteria. Of the linear systems, 86% were classified as CL, with 14% as CLB. Due to the low sample sizes associated with the CLB category, these systems have been lumped into the CL category for the statistical analysis in section 5. Nonlinear systems were led by the CNL category with 61%, followed by DA with 39%.

For the stratiform rain area classification, TS dominated with 67% of the linear systems. The TS/PS type (generally large systems since stratiform rain occurred in both regions) garnered the second highest total with 16%. The categories of LS, PS, and LS/PS (substantial areas of each) all had five or fewer occurrences in each model. Little statistical significance of LS, PS and LS/PS classifications was found, likely owing to the small sample size in each of these categories. These results were fairly similar to the Parker and Johnson (2000) survey of central U.S. linear MCS. They found TS was the dominant mode, though only accounting for 40% of the cases. The TS/PS type was second highest with 18%. In our study, the other categories of stratiform had slightly less of a representation than in the Parker and Johnson (2000) study, due to a greater domination of the TS type.

BB with 23%. EA had only one (1%) occurrence for each model CRA and was, therefore, excluded from further study. The results for BA differed greatest from Bluestein and Jain (1985) who found these events only 20% of the time for severe squall line cases in Oklahoma.

Since the classification scheme is conditioned on the observed system, forecasts that miss the event (i.e., no or very little precipitation is simulated) are included in the dataset. In totaling the number of MCS cases where no forecast rain volume existed above the 0.25 inch rainfall threshold, there were 19 (13%) cases in the Eta, 31 (22%) cases in the MM5, and 20 (14%) cases in the WRF. Thus, the MM5 had a larger number of missed events than the Eta or WRF.

5. CRA STATISTICAL ANALYSIS

Statistics were calculated for the following parameters: rain volume, rain rate, maximum grid point rainfall, phase displacement, and MSE decomposition. This analysis was performed for all of the observed systems over the Plains meeting minimum MCS criteria. Errors were then examined as a function of the observed system morphology.

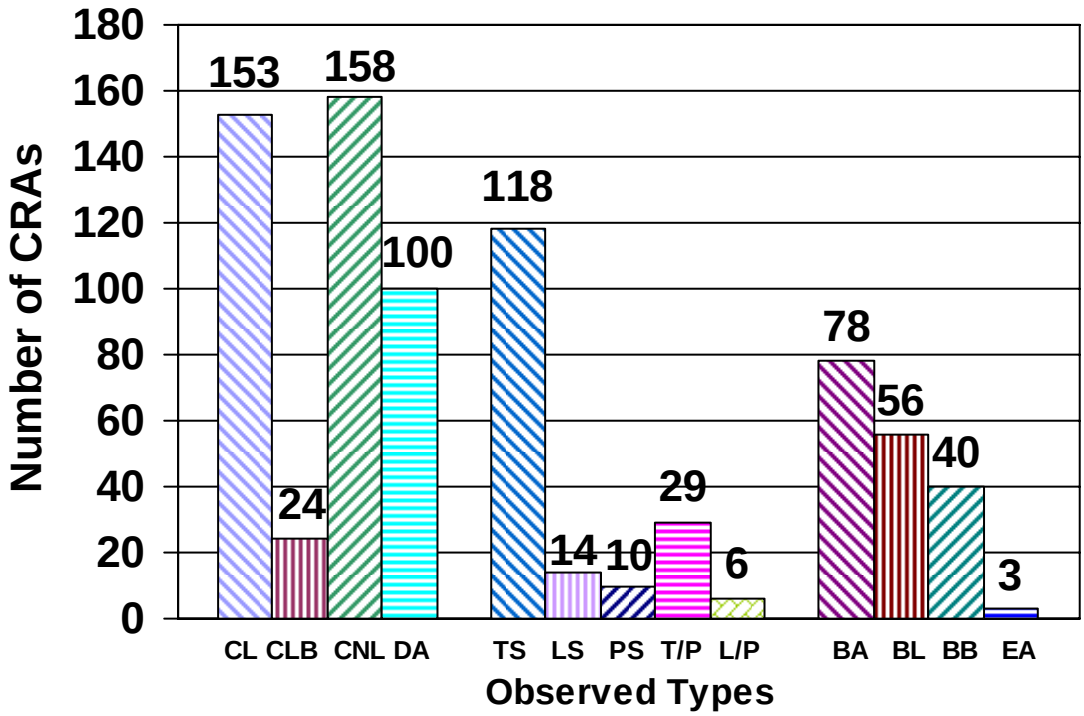


Figure 2. Histogram of observed systems (general types, stratiform types for linear systems, and development types for linear systems) for all CRAs identified by the EMT applied to the Eta, MM5, and WRF models.

Among the development types, BA was the most common with 44%, followed by BL with 32%, and then

All statistical results discussed in this section were formally evaluated by a Student's t-test, a multiple comparison analysis of variance (ANOVA) utilizing Tukey's Honestly Significant Difference (HSD) procedure (1953), and Levene's Test for Homogeneity of Population Variances (1960). These tests determined statistical significance at the .05 alpha level. For the t-test, basic assumptions were made regarding adequate sample sizes, approximate normality, and that the data comprising each sample were randomly selected from their larger population. For Tukey's HSD, assumptions of approximate normality and nearly equal variances in samples and populations were made in order to accurately perform this test, with deviations noted. Thus categories with extreme skewness or many outliers and vastly different sample sizes were excluded from discussion below. Levene's test was used to determine if population variances were not all equal for multiple comparisons.

The Student's t-test determined whether errors between forecast and observed values were biased for each type (e.g., is a mean wet bias in the DA category for the Eta truly statistically significant?). Tukey's HSD determined statistical significance of differences in mean errors between types in a given model (e.g., if both DA and CNL have a statistically significant mean wet bias for the WRF model, does one type have a greater mean bias versus the other?). This conservative test was performed to protect the true alpha level of .05 during multiple comparisons from the effects of multiplicity (e.g., Wilks 1995; Ott and Longnecker 2001). All graphical results are presented by using box plots and mean diamonds. The box represents the interquartile range, from the 25th to the 75th percentile, and the line through this box represents the median. The whiskers extend from the 25th and 75th percentiles to the outermost minimum and maximum values of the sample within 1.5 times the interquartile range. The mean diamond represents the mean (middle line) of the sample and 95% confidence intervals (apex of lines).

5.1 *Rain volume*

The Eta showed a mean dry bias with the CL category (as noted in section 4, for the CRA statistical analysis this category represents the combination of CL and CLB systems) and a mean wet bias for the DA category (Fig. 3). The mean bias of the CL category was significantly drier compared to the CNL and DA categories. This confirms that the Eta produces too little rain volume for linear systems, and this behavior differs from its performance with non-linear systems. We speculate that the mean dry bias with linear systems reflects the lack of transport of condensate away from more intense convective cells (which is not included in the BMJ convective scheme), a process known to be very important in the upscale growth of organized linear systems (e.g., Rutledge, 1986).

Both the MM5 and WRF showed a mean dry bias for all three general types (Figs. 4 and 5). The MM5 had no categories that were significantly different from the

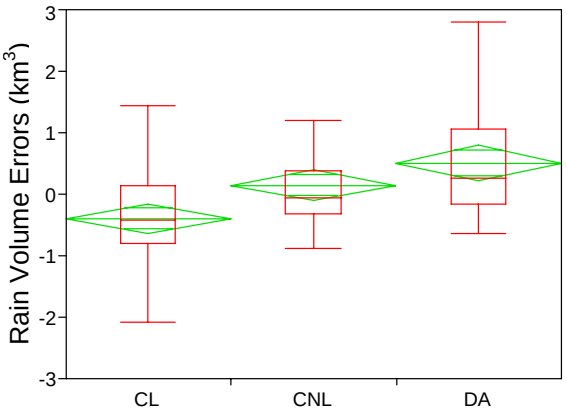


Figure 3. Box plots and mean diamonds for errors (forecast – observed) in rain volume (km^3) for general types in the Eta.

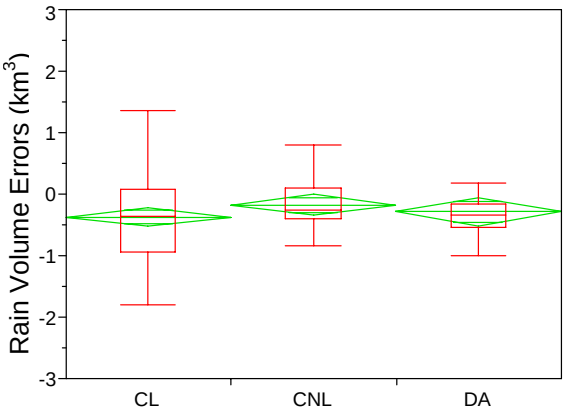


Figure 4. Same as in Fig. 3 except for the MM5.

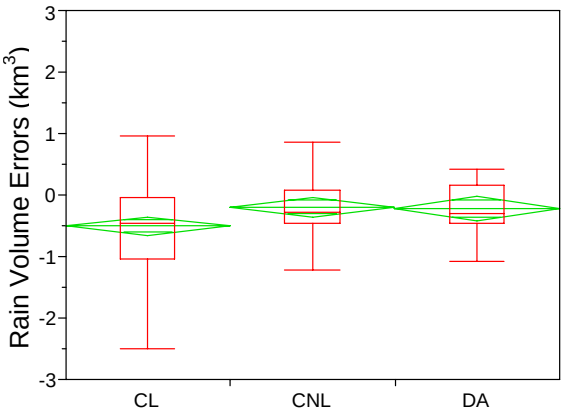


Figure 5. Same as in Fig. 3 except for the WRF.

other categories. In WRF, the CL category was significantly drier than the CNL and DA categories. It should be noted that Levene's test for the assumption of equal variances was not passed for the WRF comparison, due to the larger spread in the CL category versus the CNL and DA categories. However, as in the Eta model, the WRF appears to have larger dry biases for linear systems than non-linear ones.

As might be expected since all three models had a mean dry bias with the linear category, a dry bias was also apparent with the dominant stratiform type, TS. For development types, the Eta had a mean dry bias with both the BB and BL categories. The mean for the BB category was significantly higher than the BA category. However, the assumption of equal variances was not validated for this comparison. Dry biases were present in the MM5 and WRF for both BA and BL categories. Differences in biases were not significant among development types for the MM5 and WRF.

5.2 *Rain rate*

The Eta's forecast average rain rate (for all CRA points above the 0.25 inch threshold) was significantly lower than observed for both CL and CNL general categories. It was also significantly lower for the TS and TS/PS stratiform categories and the BB, BA, and BL development categories. In addition, the Eta produced nearly the same average rain rate for practically all general types (Fig. 6), unlike observations (Fig. 7), implying the model may not have the capability to differentiate its rate of rainfall for highly efficient precipitation systems from those with lower efficiency. Gallus (1999) showed that the Eta with the BMJ convective scheme was fairly insensitive to changes in horizontal grid resolution. He speculated that the BMJ scheme was so aggressive at drying the atmosphere that small-scale structures more likely to be produced in the grid-resolved component of the rainfall were often eliminated. Operational forecasters have long noted that the rainfall forecasts from the Eta appear to be overly smooth and lack fine-scale structure. The current analysis agrees with those observations.

However, for both the MM5 and WRF, the CL category had a significantly higher forecast average rain rate than that observed (Figs. 8 and 9). The mean errors of the CL category were also significantly higher than the DA category. For stratiform types, these same trends were noted. Both models had significantly higher average rain rates than observed in the TS category, a result consistent with a failure to develop larger areas of lighter stratiform rain (such that the heavier convective rates dominated these systems). For development types, the MM5 and WRF were both significantly higher than observed with the BB category. The MM5 and WRF results are in contrast to the much lower average rain rates of the Eta forecasts.

These results suggest a systematic rainfall distribution and amount error arising from problems with predic-

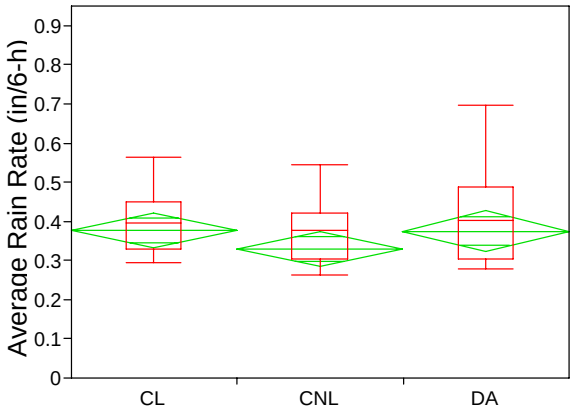


Figure 6. Box plots and mean diamonds for forecast average rain rate (in/6-h) for general types in the Eta.

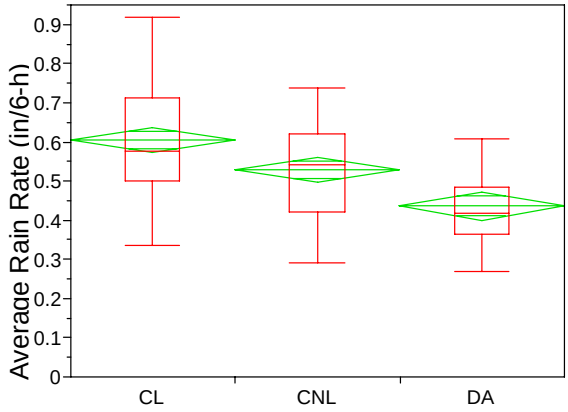


Figure 7. Same as Fig. 6 except for observed average rain rate (in/6-h).

tion of cold pool dynamics. Weisman et al. (1997) showed from three-dimensional midlatitude squall line simulations performed at a variety of grid resolutions that a delayed strengthening of the cold pool occurs with explicit models run at resolutions coarser than 4 km. Since the cold pool is crucial to the evolution of an MCS into an upshear-tilted mature system, such models can be expected to underestimate the trailing stratiform precipitation region commonly produced by the upshear-tilted front-to-rear flow, while overpredicting the precipitation in the convective leading line. Both characteristics are observed with the 12-km MM5 and WRF models in the present study.

5.3 *Maximum rainfall*

Maximum rainfall was defined as the highest observed amount of precipitation in the model's 12-km grid

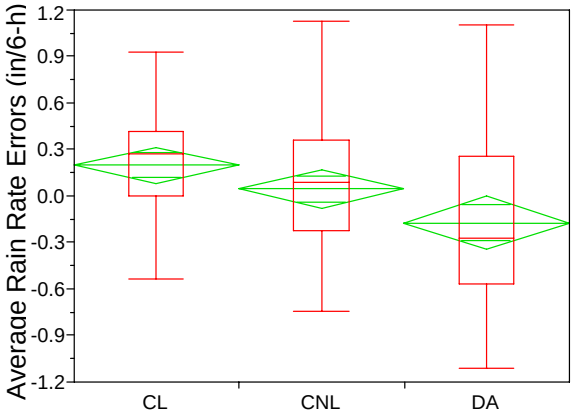


Figure 8. Box plots and mean diamonds for average rain rate errors (in/6-h) for general types in the MM5.

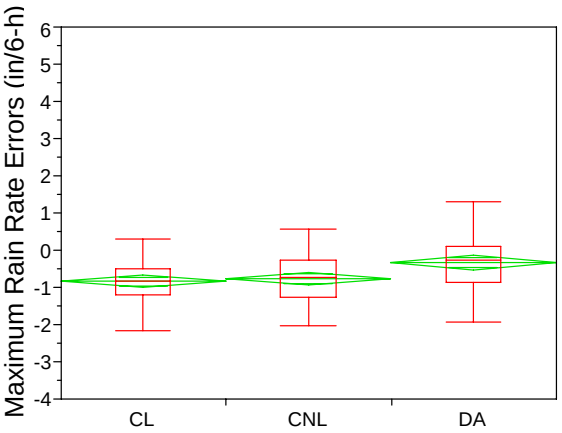


Figure 10. Box plots and mean diamonds for maximum rain rate errors (in/6-h) for general types in the Eta.

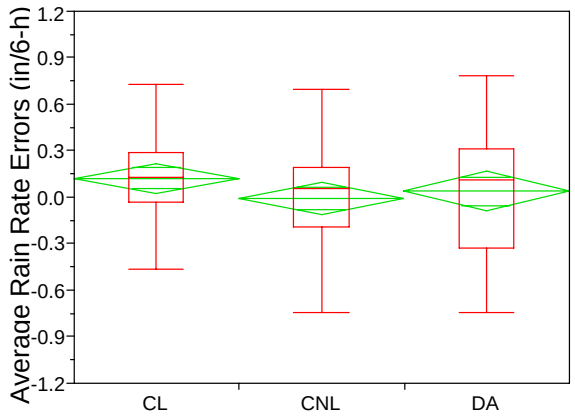


Figure 9. Same as Fig. 8 except for the WRF.

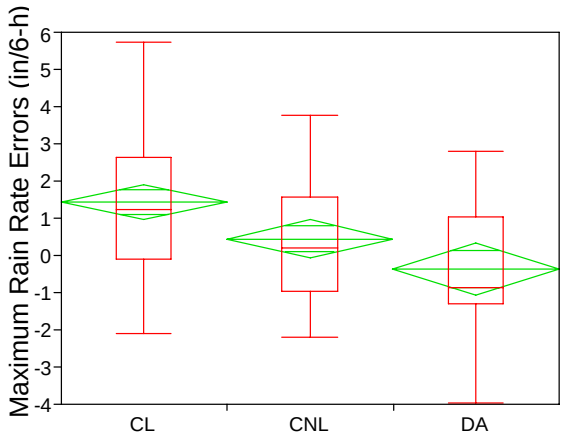


Figure 11. Same as Fig. 10 except for the MM5.

and in the filtered stage IV observed accumulation grid. The Eta significantly underpredicted average rainfall maxima overall, for all general and development types, as well as TS and TS/PS stratiform types. Both the CL and CNL categories had greater mean dry biases than the DA category (Fig. 10). For the development types, the BL category had a greater mean dry bias compared to the BA category.

As with the average rain rate category, the Eta was very uniform in its distribution of average maximum rain rate for each system type. The tendency of the Eta to have far smaller average maximum rain rates than observed agrees with Gallus (1999), who showed that the use of the BMJ scheme prevented large rainfall amounts from occurring with fine grid resolution. When the Kain-Fritsch scheme was used instead, Gallus (1999) noted that much larger rain rates resulted. He showed the maximum rain rates in simulated convective systems occurred in regions with large grid-resolved

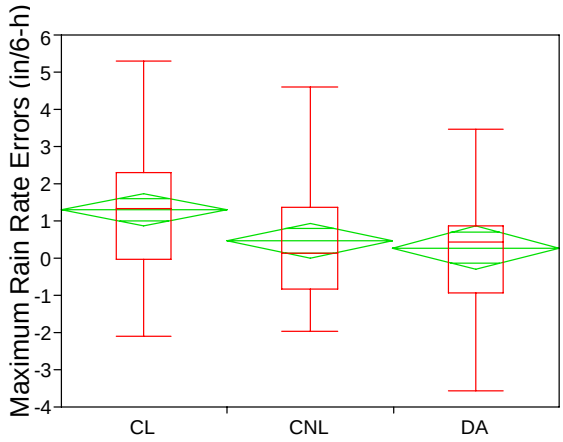


Figure 12. Same as Fig. 10 except for the WRF.

rainfall components.

Once again the MM5 and WRF results were in stark contrast to the Eta (Figs. 11 and 12). Both the MM5 and WRF had significantly larger maximum rainfall rates, on average, than observed for the CL and CNL types. For both models, CL systems had significantly larger wet biases than both the CNL and DA types. This trend continued into the stratiform categories with TS forecasts being significantly wetter than observed in both the MM5 and WRF. For all three development types, the MM5 and WRF had a mean wet bias. The MM5 and WRF also exhibited much more variability with the spread of the interquartile range (from the 0.25 to the 0.75 percentile) usually double that of the Eta for most types. Since the MM5 and WRF typically underestimated rain volume, the greater rainfall intensities are consistent with much smaller rainfall areas than observed.

5.4 *Phase displacement errors*

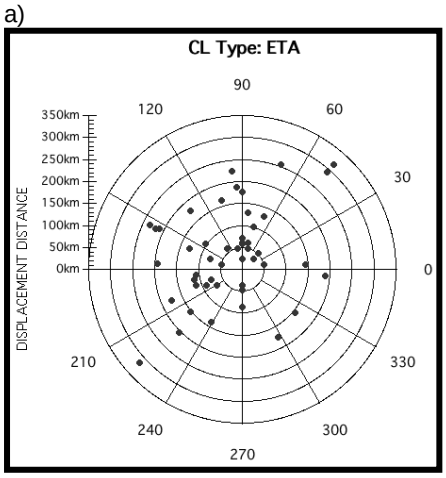
None of the models displayed a strongly preferred direction and magnitude of displacement error vectors, for any particular MCS classification except for the CL type. All three models exhibited a majority of displacements from the northwest for this type (Fig. 13). These systems were likely forecast too slowly by the three models (especially MM5 and WRF). This may suggest that MM5 and WRF simulated cold pools for squall line systems were too weak or delayed, a hypothesis fully consistent with the rainfall rate bias problems (underprediction of the stratiform rain region, overprediction of the rain rates in the convective leading lines) discussed in section 5.2. In the Eta, the BMJ convective scheme does not directly affect the model environment below the sub-cloud layer. This makes the scheme's behavior difficult to correlate to specific observed physical processes (Kain et al. 2003). Consequently, linear MCS cold pools are not realistically simulated.

5.5 *MSE decomposition*

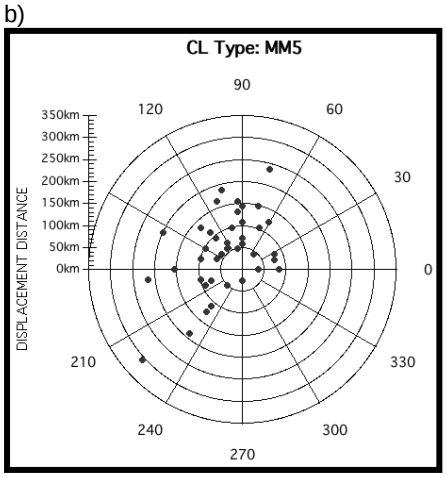
Given similar observed average rainfall volumes between the models' CRAs, one can test whether a certain model had significantly lower or higher average MSE than the others for a specific type. The average total MSE for the Eta was significantly lower than both the MM5 and WRF for the CL (Fig. 14) and CNL general types. TS was the only stratiform type to be significantly lower in the Eta versus MM5 and WRF. However, the test for assumption of equal variances in both the CL and TS types failed, since both the MM5 and WRF clearly had much larger variances than the Eta.

For the CL and CNL types, both the MM5 and WRF had their largest source of total MSE from pattern errors, followed by displacement errors, and then volume errors. The Eta was similar in this distribution for the CL type. But for the CNL type, larger errors for pattern were

a)



b)



c)

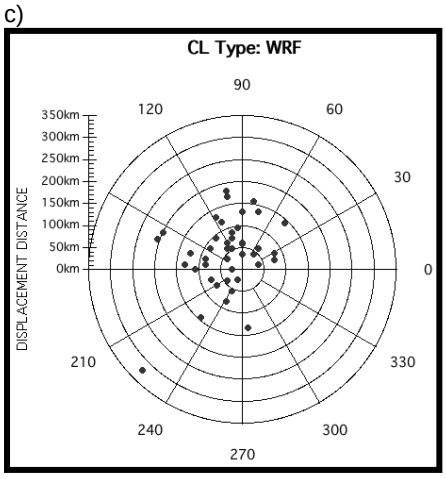


Figure 13. Phase displacement errors for the CL general type in the Eta (a), MM5 (b), and WRF (c) models. Dots represent direction and magnitude of displacement errors from original to shifted forecast. Distribution is only shown for those systems which had a displacement calculated.

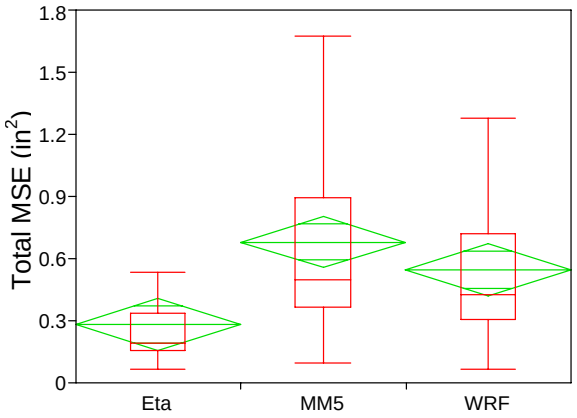


Figure 14. Box plots and mean diamonds for total MSE (in^2) for the CL type in the Eta, MM5, and WRF.

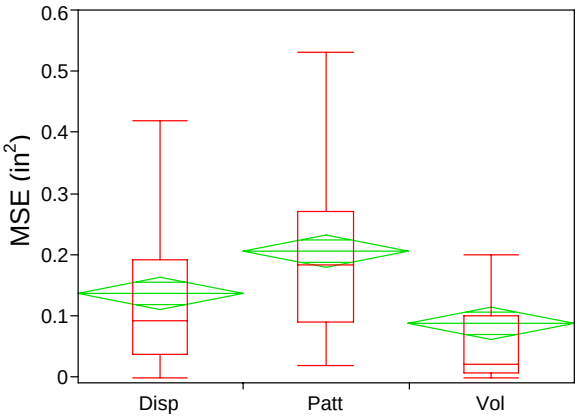


Figure 15. Box plots and mean diamonds for displacement, pattern, and volume errors (in^2) for the CL type with combined results from the Eta, MM5, and WRF.

followed by volume errors, and then displacement errors. The DA type did not display the same distribution as the other three general types; there were no significant differences between the types of errors for any of the models. The magnitude of errors was nearly equally distributed among all decomposition terms for this type.

Pattern error was the largest source of average error in the MSE decomposition for all three models for CL and CNL types. The Eta model was also significantly lower in average total MSE for these categories compared to the MM5 and WRF. As mentioned in section 2g, pattern error is strongly influenced by variability, which is a function of the effective model resolution. Since the Eta model replicates scales 3 to 4 times larger than the MM5 and WRF, the Eta has lower variability. All other factors being equal, a model with lower variability will have lower MSE. The magnitude of pattern errors are around twice as large for the MM5 and WRF compared to the Eta in the CL type (Fig. 15), similar to the magnitude of total MSE. Operational models have tended to be designed to produce smoothly varying QPFs, despite the preference of some human forecasters for more realistic-looking detail and the increasing simulation by research models of finer representations of QPF.

6. CONCLUSIONS

The EMT was modified to optimize detection of MCSs occurring over the central U.S. and applied to forecasts of convective system rainfall from the 12-km Eta, MM5, and WRF models during IHOP 2002. This technique allowed for the determination of errors as a function of observed convective system morphology, a procedure not possible with typical gridpoint-to-gridpoint domain-wide verification. No attempts were made to use the EMT for determining errors as a function of forecast convective system morphology, due to an inability of 12-km models to properly simulate detailed convective system characteristics.

Systematic deficiencies were found in these models for various types of convective systems, when using the error measures supplied by the EMT. While almost all of the differences found in comparing the Eta and the MM5/WRF were not surprising, the results as a function of the observed convective system morphology provide additional insight into the spectrum of MCS errors in each model. These results should help modelers in their assessments and may have some limited relevance to forecasters. For modelers, the error metrics can point out certain morphological types where the model has a systematic bias or relatively inaccurate forecast compared to other observed types. For forecasters, the utility of these results depends ultimately on an a priori knowledge of likely convective system morphological evolution, based on conceptual/numerical models and experience. Knowing what the numerical model QPFs typically depict for a certain type of system, forecasters can further confirm or reject their forecast formulated on the environmental wind/thermodynamic fields and other observations. However, forecasting warm-season convective system morphology is in itself a problematic and uncertain process. Thus, the conditional verification information provided here will probably be of more use as an assessment for modelers, rather than as a forecasting tool.

The modified EMT suggested that the Eta underestimated rain volume for linear systems and overestimated it for discontinuous non-linear ones, while both the MM5 and WRF underestimated volume for all systems. The Eta also produced average rain rates and peak rainfall amounts that were much too light for almost all systems, likely due to its typically low-variability and overly smoothed QPFs. On the other hand, the MM5 and WRF both produced average rain rates and

peak rainfall amounts that were higher than observed for most linear classifications. These two models were dry-biased with rain volume reflecting a large underestimate of areal coverage compared to observations for linear systems. All three models forecast rainfall too far northwest for linear systems. These results suggest a systematic rainfall distribution and amount error arising from problems with prediction of cold pool dynamics, following Weisman et al. (1997). The Eta had smaller total mean square errors than the MM5 and WRF for both CL and CNL systems, as well as for TS types. The smaller errors again likely reflect its tendency to produce smoother rainfall fields than the WRF and MM5. For all general MCS types (except DA), the largest contributors to total MSE were pattern errors, typically followed by displacement, and then volume errors.

Overall, the modified EMT suggested various systematic errors are dependent on convective system type and model. No one general type or model was consistently better or worse than the other types. Out of the stratiform types, TS systems typically had the same biases as those of CL systems. Due at least partly to small sample sizes, almost all of the other stratiform types were not found to have significant biases. Error measures did not consistently differ among the development types. It is plausible that processes occurring during development operate on scales too small for a 12-km model to differentiate.

In future work, this technique and observed morphology classification scheme could be used to evaluate other models or different versions of the same model. The EMT could also potentially be applied to verifying human forecasts, in addition to those of numerical models. Since the National Weather Service has moved into the digital forecast era with the National Digital Forecast Database (Glahn and Ruth 2003), an object-oriented gridded verification could occur between human forecasts and numerical models. Questions such as how does overall rainfall volume and rate differ from the human versus model forecast and do human forecasts exhibit the same northwest bias for linear MCSs could be answered.

The EMT's flexibility for user-defined parameters in object-oriented verification, along with its production of several error metrics at once, makes the technique a valuable tool in the assessment of forecasts. By developing a classification scheme based upon the observed morphology, the technique can further differentiate its error measures, and provide modelers with error information for specific types of observed systems. Such information may be useful in pinpointing specific shortcomings in model physics or dynamics, allowing for more potential improvement in numerical forecasts.

7. ACKNOWLEDGMENTS

This research was supported in part by NSF Grant 0226059 and by NOAA support to the Forecast Systems Laboratory provided through the U.S. Weather Research Program.

8. REFERENCES

Albers, S., J. McGinley, D. Birkenheuer, and J. Smart, 1996: The Local Analysis and Prediction System (LAPS): Analysis of clouds, precipitation, and temperature. *Wea. Forecasting,* **11**, 273-287.

Baldwin, M. E., and M. Wandishin, 2002: Determining the resolved spatial scales of Eta model precipitation forecasts. *Preprints, 15th Conf. on Numerical Weather Prediction*, San Antonio, TX, Amer. Meteor. Soc., 85-88.

Bluestein, H. B., and M. H. Jain, 1985: Formation of mesoscale lines of precipitation: severe squall lines in Oklahoma during the spring. *J. Atmos. Sci.,* **42**, 1711-1732.

Bullock, R., B. G. Brown, C. A. Davis, K. W. Manning, and M. Chapman, 2004: An object-oriented approach to quantitative precipitation forecasts. *Extended Abstracts, 17th Conf. on Probability and Statistics in the Atmospheric Sciences*, Seattle, WA, Amer. Meteor. Soc., J12.4. [Available online at http://ams.confex.com/ams/84Annual/techprogram/session_16618.htm].

Du, J., S. L. Mullen, and F. Sanders, 2000: Removal of distortion error from an ensemble forecast. *Mon. Wea. Rev.,* **128**, 3347-3351.

Duchon, C. E., 1979: Lanczos filtering in one and two dimensions. *J. Appl. Meteor*., **18**, 1016–1022.

Ebert, E. E., and J. L. McBride, 2000: Verification of precipitation in weather systems: determination of systematic errors. *J. Hydro.,* **239**, 179-202.

Gallus, W. A., Jr., 1999: Eta simulations of three extreme precipitation events: Impact of resolution and choice of convective parameterization. *Wea. Forecasting*, **14**, 405-426.

Glahn, H. R., and D. P. Ruth, 2003: The new digital forecast database of the National Weather Service. *Bull. Amer. Meteor. Soc.*, **84**, 195-201.

Grell, G. A., J. Dudhia, and D.S. Stauffer, 1995: A description of the fifth-generation Penn State-NCAR mesoscale model (MM5). NCAR Technical Note, NCAR/TN-398 + STR, 122 pp.

Harris, D., E. Foufoula-Georgiou, K. K. Droegemeier, and J. J. Levit, 2001: Multiscale statistical properties of a high-resolution precipitation forecast. *J. Hydrometeor*., **2**, 406-418.

Hoffman, R.N., Z. Liu, J.-F. Louis, and C. Grassotti, 1995: Distortion representation of forecast errors. *Mon. Wea. Rev.,* **123**, 2758-2770.

Janjic, Z. I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927-945.

Kain, J. S., M. E. Baldwin, P. R. Janish, S. J. Weiss, M. P. Kay, and G. W. Carbin, 2003: Subjective verification of numerical models as a component of a broader interaction between research and operations. *Wea. Forecasting*, **18**, 847-860.

Koch, S. E., and Coauthors, 2004: Real-time applications of the WRF model at the Forecast Systems Laboratory. *Preprints, 20th Conf. on Weather Analysis and Forecasting and 16th Conf. on Numerical Weather Prediction,* Seattle, WA, Amer. Meteor. Soc., 12.1. [Available online at http://ams.confex.com/ams/84Annual/techprogram/session_16578.htm].

Levene, H., 1960: Robust tests for the equality of variances. *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, I. Olkin et. al., Eds., Stanford University Press, 278-292.

McGinley, J. A., and J. R. Smart, 2001: On providing a cloud-balanced initial condition for diabatic initialization. *Preprints, 18th Conf. on Weather Analysis and Forecasting*, Ft. Lauderdale, FL, Amer. Meteor. Soc.

Murphy, A. H., 1995: The coefficients of correlation and determination as measures of performance in forecast verification. *Wea. Forecasting,* **10**, 681-688.

Ott, R. L., and M. Longnecker, 2001: *An Introduction to Statistical Methods and Data Analysis.* Duxbury, 1152 pp.

Parker, M. D., and R. H. Johnson, 2000: Organizational modes of midlatitude mesoscale convective systems. *Mon. Wea. Rev.,* **128**, 3413-3436.

Rogers, E., and Coauthors, 1998: Changes to the NCEP operational “early” Eta analysis/forecast system. NWS Tech. Procedures Bull. 477, National Oceanic and Atmospheric Administration/ National Weather Service, 14 pp. [Available from Office of Meteorology, National Weather Service, 1325 East–West Highway, Silver Spring, MD 20910.].

Rutledge, S. A, 1986: A diagnostic numerical study of the stratiform region associated with a tropical squall line. *J. Atmos. Sci.*, **43**, 1337-1358.

Skamarock, W. C., J. B. Klemp, and J. Dudhia, 2001: Prototypes for the WRF (Weather Research and Forecasting) model. Preprints, *9th Conf. on Mesoscale Processes,* Fort Lauderdale, FL, Amer. Meteor. Soc., J11-J15.

Tukey, J. W., 1953: The problem of multiple comparisons. *The Collected Works of John W. Tukey Volume VIII, Multiple Comparisons: 1948-1983.* H.I. Braun, Ed., Chapman Hall, 1-300.

Weisman, M. L., W. C. Skamarock, and J. B. Klemp, 1997: The resolution dependence of explicitly modeled convective systems. *Mon. Wea. Rev.,* **125**, 527-548.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.

Zepeda-Arce, J., E. Foufoula-Georgiou, and K. K. Droegemeier, 2000: Space-time rainfall organization and its role in validating quantitative precipitation forecasts. *J. Geophys. Res*., **105**, 129-146.